FEBS Letters 589 (2015) 342–348

journal homepage: www.FEBSLetters.org

Botulinum neurotoxin homologs in non-*Clostridium* species

Michael J. Mansfield, Jeremy B. Adams, Andrew C. Doxey *

Department of Biology, University of Waterloo, 200 University Ave. West, Waterloo, Ontario N2L 3G1, Canada

ARTICLE INFO

Article history:
Received 15 October 2014
Revised 2 December 2014
Accepted 9 December 2014
Available online 23 December 2014

Edited by Takashi Gojobori

Keywords:
Botulinum neurotoxin
Clostridial neurotoxin
Toxin
Lateral gene transfer
Clostridium
Weissella

ABSTRACT

Clostridial neurotoxins (CNTs) are the deadliest toxins known and the causative agents of botulism and tetanus. Despite their structural and functional complexity, no CNT homologs are currently known outside *Clostridium*. Here, we report the first homologs of *Clostridium* CNTs within the genome of the rice fermentation organism *Weissella oryzae* SG25. One gene in *W. oryzae* S25 encodes a protein with a four-domain architecture and HExxH protease motif common to botulinum neurotoxins (BoNTs). An adjacent gene with partial similarity to CNTs is also present, and both genes seem to have been laterally transferred into the *W. oryzae* genome from an unknown source. Identification of mobile, CNT-related genes outside of *Clostridium* has implications for our understanding of the evolution of this important toxin family.

1. Introduction

Clostridial neurotoxins (CNTs), which include the botulinum neurotoxins (BoNTs) and tetanus neurotoxin (TeNT), are the most deadly toxins known and the causative agents of botulism and tetanus neuroparalytic diseases [1]. The toxins are present in four phylogenetically distinct lineages of *Clostridium botulinum* as well as some strains of *Clostridium butyricum* and *Clostridium baratii* [2], and are encoded within conserved gene clusters that undergo transfer and recombination between strains [3]. Despite being a subject of intense study and the wealth of structural and functional information on CNTs [4], little is known regarding their evolutionary origins and relationships to other proteins.

CNTs have a complex multidomain architecture composed of an N-terminal catalytic peptidase domain followed by a translocase domain and two receptor binding domains (N- and C-terminal binding domains, respectively). Each domain plays a unique role in the functional specificity of BoNTs and TeNTs towards their neuronal targets. Following translation, the peptidase domain (light chain) is cleaved from and disulfide linked to the remaining protein (heavy chain) [5]. The two receptor binding domains (PFAM IDs Toxin_R_bind_N and Toxin_R_bind_C) direct toxin complexes to neurons and enter the cell via receptor-mediated endocytosis. At this point, the translocase domain (PFAM ID Toxin_trans) induces endosomal pore formation, and the peptidase domain is inserted into the cytosol. The peptidase (PFAM ID Peptidase_M27), mediated in part by a conserved zinc peptidase motif (HExxH), serves to cleave neuronal SNAREs (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors), preventing exocytosis of acetylcholine and resulting in flaccid paralysis [5]. Tetanus neurotoxins have an identical domain structure and employ a similar mechanism, though are transported retroaxonally after translocation and instead result in spastic paralysis [6].

Non-toxic non-hemagglutinins (NTNHs), paralogs of BoNTs, are genomically located adjacent to the BoNT gene in all known non-tetanus gene clusters. NTNH does not insert itself into neurons, and instead forms progenitor toxin complexes with BoNTs [7]. The NTNH domain architecture also consists of a peptidase, translocase and N-terminal binding domains; however, its C-terminal binding domain, while structurally related to Toxin_R_bind_C, is quite distinct in sequence. Additionally, its peptidase domain lacks the catalytic HExxH motif found in BoNTs.

Given the structural and mechanistic complexity of CNTs as well as their limited phyletic occurrence in *Clostridium*, the question of how and why these proteins have evolved is a fascinating one. In previous work [8], we searched for homologs of CNTs in existing genomes. Some intriguing local similarities between genes within the *C. botulinum* strain A Hall neurotoxin gene cluster were

Abbreviations: CNT, clostridial neurotoxin; BoNT, botulinum neurotoxin; TeNT, tetanus neurotoxin; NTNH, non-toxic non-hemagglutinin; ORF, open reading frame

* Corresponding author.
E-mail address: acdoxey@uwaterloo.ca (A.C. Doxey).

http://dx.doi.org/10.1016/j.febslet.2014.12.018

identified, suggesting that the toxin may have originated through repeats and rearrangements of neighboring genes – the most obvious case being an apparent ancestral tandem duplication that gave rise to BoNT and NTNH. However, no homologs of an entire CNT gene could be detected outside of *C. botulinum* or its phages. Thus, whether CNTs originated within or outside of the *Clostridium* lineage is still unclear.

Here, we have repeated our searches for CNT homologs in far greater breadth, including many metagenomes and expanded public databases. We report the first CNT homologs outside of *Clostridium*, thus expanding the family of CNTs, and our understanding of CNT evolution.

2. Results and discussion

Using BLAST [9] and HMMER [10], we searched genomic and metagenomic datasets for multi-domain homologs of clostridial neurotoxins, which have not been identified outside of *Clostridium*. Despite the ubiquity of *C. botulinum* in natural environments such as soil, a HMMER search of 2829 public metagenomes via the IMG/M database revealed no homologs containing multiple neurotoxin domains.

However, we detected two proteins (ORF1, NCBI gi: 653854119; and ORF2, NCBI gi: 653854116) from the recently sequenced *Weissella oryzae* SG25 genome [11], as possessing multi-domain homology to CNTs (Fig. 1). Reciprocal homology, which is strong support for an evolutionary relationship, was detected in all directions between representative CNTs and ORF1/ORF2 through single or multiple BLAST iterations (PSI-BLAST) with significant *E*-values (Fig. 1A). Homology was also detected through pHMMER searches [10] (data not shown).

W. oryzae is a gram-positive, non-motile, non-spore-forming, facultative anaerobe [13]. Like *Clostridium*, it is a member of the phylum *Firmicutes*, though it belongs within the separate class *Bacilli*, order *Lactobacillales*, and family *Leuconostocaceae*. The species is most closely related to *Weissella soli* (97% 16S identity), and shares only ~83% 16S identity with *C. botulinum* strains. *W. oryzae* was initially isolated from fermented rice grains [13]; fermented foods are an environment known to harbour *C. botulinum* [14]. Other members of *Weissella* have been discovered in similar food fermentation and soil environments [15–17]. Additionally, the detected *Weissella* genes are located immediately adjacent to each other in the *W. oryzae* genome (Fig. 1B), a gene architecture reminiscent of the BoNT and NTNH genes in *C. botulinum*.

PFAM analysis shows that ORF1 and ORF2 have domains corresponding to those in CNTs (Fig. 1D). ORF1 has a four-domain architecture identical to that in BoNTs, while ORF2 possesses only the N-terminal peptidase M27 domain ($E = 1e{-}25$) and a partial match to the translocase domain (440–495 aa, $E = 3.2e{-}4$) (Fig. 1D). In place of the N and C binding domains, ORF2 contains two bacterial immunoglobulin-like ("Big 3") domains in this region, which may play an analogous function given their similarly beta-rich structure and association with carbohydrate-binding [18].

To assess the extent to which homology could be detected across the full length proteins, a sliding window BLAST analysis was performed. Here, ORF1 and ORF2 query sequences were divided into 100 amino acid segments, each of which was separately searched against full length CNTs (Fig. 1C). Extended ORF1 homology to BoNT was detected across all four domains as suggested by the diagonal pattern in the dot plot shown in Fig. 1C. ORF1 shows substantially less similarity to NTNH A1. ORF2 on the other hand shows more similarity to NTNH A1 than BoNT A1, mainly localized to the peptidase region. These patterns are therefore consistent with the domain predictions (Fig. 1D).

We then aligned representative CNT sequences from each toxinotype with ORF1 and ORF2 (Fig. 1E; for the full length alignment, see Supplementary information). The alignment itself reveals clear homology, as is evident for a subregion of the peptidase domain (Fig. 1E). Moreover, the critical, conserved HExxH zinc metallopeptidase motif of BoNTs and TeNT is also shared in the BoNT-like ORF1 (Fig. 1E). The NTNH group lacks this motif, as does ORF2. Both ORF1 and ORF2, along with NTNH, lack a conserved PYxGxAL motif that in BoNTs functions in toxin translocation [19].

Since mosaicism has been observed in BoNT C/D and D/C [20,3], we separately compared the heavy and light chain regions of ORF1/2 to known BoNTs (Supplementary information). This revealed the light chain to possess greater average amino acid sequence identity to BoNTs (18.7%) than the heavy chain (15.9%). Furthermore, both regions of ORF1 were most similar to BoNT/E1 and BoNT/F1. Thus, no evidence of mosaicism was observed.

Maximum-likelihood (ML) and neighbor-joining (NJ) phylogenetic analyses were then performed including representative CNTs and the ORF1 and ORF2 sequences (Fig. 2). Prior to phylogenetic analysis, the alignment was modified to remove the non-homologous C-terminal region of ORF2. The ML and NJ trees are largely congruent overall (see Supplementary information). Furthermore, the ML tree topology is similar to a recently published BoNT phylogeny [19], and BoNT clades correspond highly to their respective serotype labels. This clade-serotype association is weaker for the NTNH clade. The placement of ORF1 and ORF2 in this tree is of particular interest. Both methods reliably placed ORF1 and ORF2 as distinct, early diverging lineages, with ORF1 as sister to the BoNT clade and ORF2 as sister to the NTNH clade (Fig. 2). This topology is dependent, however, on the midpoint-estimated root position. ORF1 and ORF2 are thus related to the BoNTs and NTNHs, but according to the tree have diverged from the ancestor of these groups before they expanded in *Clostridium*.

Referring back to the genomic context, in *C. botulinum*, the 5′-NTNH-BoNT-3′ orientation is conserved among all known gene clusters, whereas the reversed 5′-ORF1 (BoNT-like) – ORF2 (NTNH-like)-3′ orientation is observed in the *W. oryzae* S25 genome (Fig. 1B). Further examination of the *W. oryzae* gene cluster also revealed no hallmarks of CNT-like gene clusters such as the botR gene [2]. Thus, the ORF1/ORF2 gene cluster is distinct from neurotoxin gene clusters in *C. botulinum*.

To further compare the ORF1/ORF2 containing contig (NZ_BAWR01000013.1) against other available genomes, we examined the top non-self BLAST matches for the 75 predicted ORFs encoded on this contig (Table 1). The most common genus identified through top BLAST matches was the closely related *Leuconostoc*, which was also one of the most frequent genera detected for all other *W. oryzae* genes (Table 1). This suggests that the contig is chromosomal and indeed derived from *Weissella* or a closely related *Leuconostocaceae* species.

Interestingly, the contig aligned relatively poorly to other existing *Weissella* genomes, containing few blocks of orthologous sequences, but aligned well to a genomic region from *Leuconostoc mesenteroides* (Fig. 3A), particularly in a large ~4 kB region 3′ to ORF2. One interesting exception, however, is a single topoisomerase IA gene (YP_004725636) from the closely related *Weissella koreensis* [13], which aligned to sequences immediately flanking ORF1 and ORF2. Although this gene is intact in *W. koreensis*, it appears to have been disrupted and partially duplicated through an insertion of the segment encoding ORF1/ORF2 in *W. oryzae* (Fig. 3B). Sequence analysis revealed the presence of a 5′ frameshift mutation in the topoisomerase IA gene and a 3′ premature stop codon, which indicates that it is likely a pseudogene. Given the lack of ORF1/ORF2 in other *Weissella* genomes, this is likely a unique insertion of an ORF1/ORF2 gene cassette into an ancestral *W. oryzae* genome.

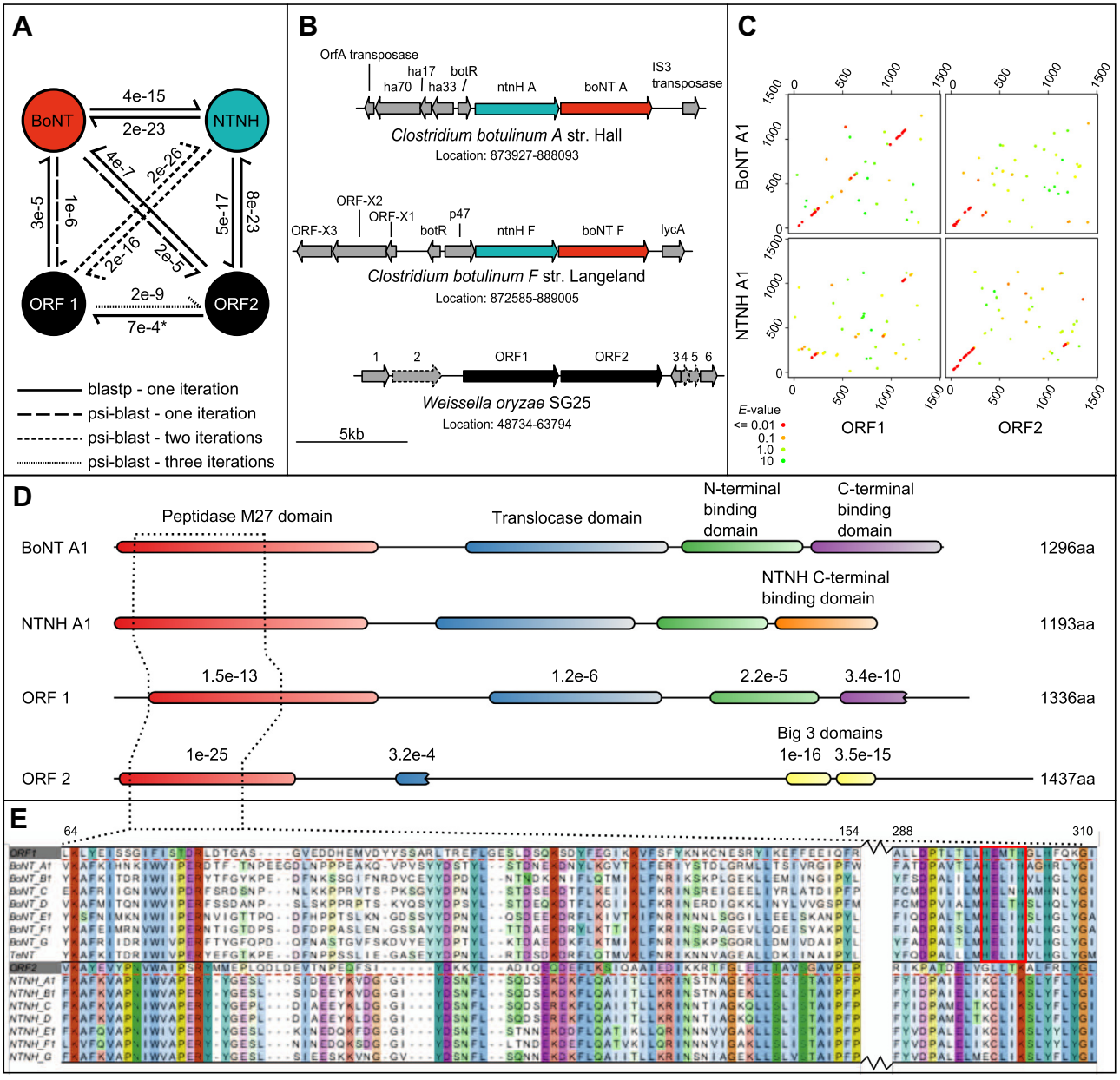

Fig. 1. Detection and sequence analysis of CNT homologs in *Weissella oryzae* S25. (A) BLAST results between CNTs and *Weissella* ORF1 (gi = 653854119) and ORF2 (gi = 653854116). Numbers are *E*-values, and arrows represent which protein was the query and which was the target. *The *E*-value to detect ORF1 from ORF2 is based on a modified ORF2 sequence (residues 1–495) with the C-terminal domains removed, since ORF2 contains common Big 3 domains that dilute the PSI-BLAST scoring matrix. (B) Gene architecture of BoNT gene clusters in two *C. botulinum* strains compared to the ORF1/ORF2 gene cluster in *W. oryzae*. Strains Hall and Langeland were selected to represent typical HA and ORF-X type gene cluster architectures, respectively. The genes surrounding ORF1 and ORF2 are depicted in (B) and numbered 1–6. 1: no putative conserved domains; 2: interrupted DNA topoisomerase IA; 3: conserved hypothetical; 4: putative DNA topoisomerase; 5: putative DNA topoisomerase; 6: conserved hypothetical. Gene coordinates were retrieved using the NCBI genome browser, and visualized using the genoplotR package for R [12]. (C) Dotplot BLAST analysis of protein similarity between BoNT, NTNH, and *Weissella* ORF1 and ORF2. Query proteins (ORF1 and ORF2) were divided into fragments of length 100 amino acids, and each fragment was separately BLAST searched against the full length target proteins (BoNT/A1 and NTNH/A1). The position and *E*-value of the top hit for each fragment is indicated in the dot plot. (D) Domain architecture of BoNT, NTNH and the two *Weissella* ORFs. The HMM alignment hit range was mapped to scale onto the CNTs and the target hit range mapped onto the *Weissella* sequences. Hits with an *E*-value greater than 1.0 have not been reported. *Note:* Although a match to the NTNH C domain was detected in ORF1 and BoNT (3.5e−5 and 0.17), predicted matches to the BoNT C domain were significantly stronger (3.4e−10 and 4.8e−75). (E) Representative alignment of selected CNTs and *Weissella* sequences. It is important to note that the HExxH peptidase motif has been conserved in both the BoNTs and ORF1, while it has diverged in the NTNHs and ORF2. For the full length alignment, see Supplementary information.

This scenario is highly reminiscent of the neurotoxin gene cluster chromosomal insertions observed in *C. botulinum* strains. Three different sites of chromosomal insertion have been observed thus far in *C. botulinum* group I (produce type A, B, F, and H neurotoxins), and two have been observed in group II, which produce type B, E or F neurotoxins (see Carter and Peck [21] for more information). In group II, neurotoxin gene cluster insertions into the *rarA* gene [20] and *topB* (topoisomerase) gene [22] have occurred, whereby both genes are apparently split by the inserted DNA. The observation here of another CNT-related gene insertion into a topoisomerase, indicates yet another link with CNT gene clusters and perhaps particularly those of group II.

Despite the homology with CNTs, the function of the *W. oryzae* ORFs is unknown and is subject for speculation. The homology,

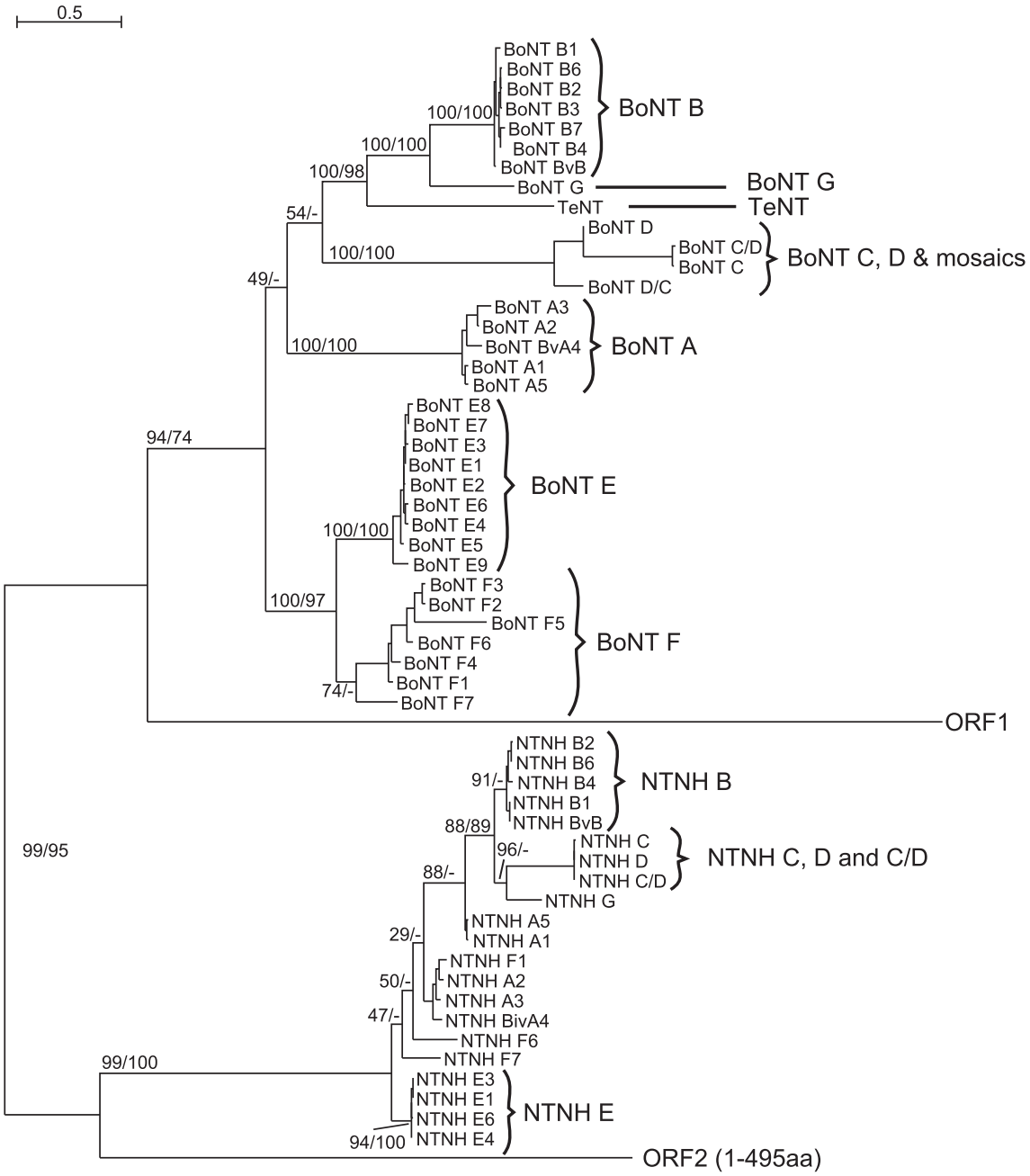

Fig. 2. Maximum likelihood phylogeny of CNTs and *W. oryzae* S25 ORF1 and ORF2 sequences. Maximum likelihood (ML) and neighbor-joining (NJ) bootstrap values are depicted above the nodes in the form (ML/NJ). Phylogenies were constructed based on the alignment in Supplementary information, with a truncated ORF2 sequence (see Section 4). The tree reveals BoNT and NTNH clades, with the BoNT clade exhibiting a topology similar to that observed previously [19], as well as two early diverging ORF1 and ORF2 lineages. ORF1 and ORF2 appear to be sister groups of BoNT and NTNH.

Table 1
Top BLAST hits for *W. oryzae* ORF1/ORF2-containing contig compared to the rest of the genome.

Rank	ORF1/ORF2 contig		Remaining genome	
	Genus	Frequency	Genus	Frequency
1	*Leuconostoc*	13	*Weissella*	867
2	*Weissella*	4	*Leuconostoc*	167
3	*Lactobacillus*	3	*Lactobacillus*	133
4 (tie)	*Lactococcus*, *Enterococcus*, *Streptococcus*	2	*Lactococcus*	111

Taxonomic composition of proteins encoded on the *W. oryzae* S25 ORF1/ORF2-containing contig (NZ_BAWR01000013.1) and remaining genome. 75 and 1836 proteins encoded on each genomic portion, respectively, were BLAST searched against the NCBI nr database, and the genus of the top significant ($E < 0.01$) non-self hit was reported. The most frequent genera are listed in the table. ORF1 and ORF2 significantly detected BoNT (top hit) and NTNH (second top hit) genes respectively from *C. botulinum*, which results in a frequency of 1 for *Clostridium*.

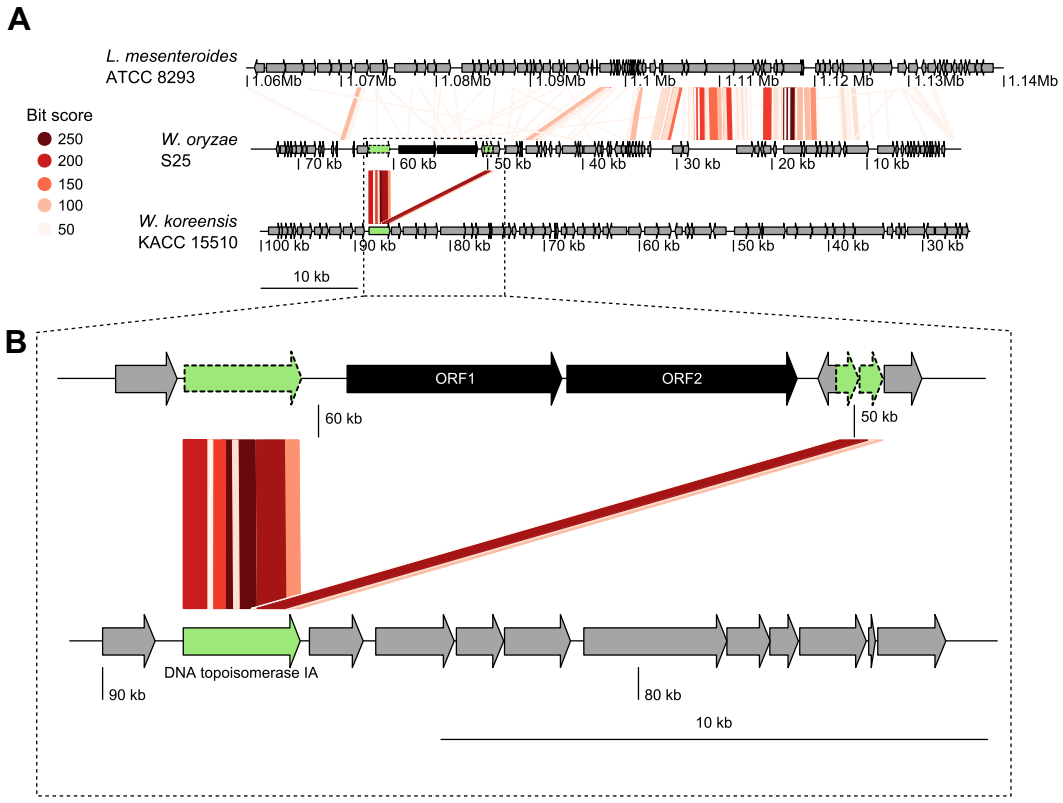

Fig. 3. Syntenic blocks between the *W. oryzae* S25 contig containing ORF1/2 and genomic regions from *W. koreensis* KACC 15510 and *L. mesenteroides* ATCC 8293. (A) Regions of similarity detected in pairwise tblastx searches are depicted as lines, coloured according to bitscore. A genomic region 3′ to ORF2 spanning roughly 4 kilobases displays homology and synteny to a genomic region from the related species, *L. mesenteroides*. (B) The genomic region immediately flanking ORF1/ORF2 is homologous to a single topoisomerase IA (YP_004725636) gene present in the *W. koreensis* genome. Sequence analysis suggests an insertion of ORF1/ORF2 into a *Weissella oryzae* topoisomerase gene, and possible pseudogenization of this gene as indicated by the presence of frameshift mutations and premature stop codons. For (A) and (B), pairwise tblastx searches were performed of *W. oryzae* contig (NCBI ID NZ_BAWR01000013.1) versus the *L. mesenteroides* ATCC 8293 (NC_008531.1) and *Weissella koreensis* KACC 15510 (NC_015759.1) genomes. Regions of similarity were visualized using genoplotR [12].

similarity in domain architecture and presence of the key HExxH motif in ORF1 all raise the possibility that ORF1 in particular performs a molecular function analogous to CNTs. Although the additional components of CNT gene clusters are lacking (botR, hemagglutinin genes, and ORF-X proteins), BoNT alone exhibits toxicity [23], and so it is plausible that these proteins function as proteolytic toxins or virulence factors. The genus *Weissella* is indeed known to include pathogens of humans and fish [24,25], but another intriguing possibility is that these CNT homologs target plants given that *Weissella* species and *Leuconostocaceae* are associated with plant degradation [26]. The possibility that there exist plant-targeting CNT homologs has been raised previously [27] since plants also possess SNAREs that mediate plant defense [28]. Ultimately, biochemical characterization of these proteins will be necessary to decipher their biological role and functional links with CNTs.

3. Conclusion

Two proteins encoded within the recently sequenced *W. oryzae* SG25 genome were detected as the first multi-domain homologs of clostridial neurotoxins. ORF1 shares homology with CNTs over its full length sequence and contains all four CNT domains. Given the divergence of ORF1/ORF2 from other CNTs and their unique phylogenetic placement within the tree (Fig. 2), they appear to represent new lineages of CNT homologs.

Since ORF1/ORF2 appears unique to *W. oryzae* S25, they have likely originated by lateral transfer and subsequent insertion into the *W. oryzae* topoisomerase IA gene from an unknown source. Based on phylogenetic analysis of ORF1 and ORF2 themselves (Fig. 2) and gene cluster differences (Fig. 1B), the donor organism is probably distinct from currently known *C. botulinum* strains. Given the mobility of this gene cassette along with its phylogenetic novelty, this finding has important implications for our understanding of CNT evolution.

4. Methods

4.1. Homology search and domain analysis

Searches were performed using the NCBI blastp web service (Refseq protein database, BLOSUM62 matrix, gap existence: 11, gap extension: 1, expect threshold 10 with conditional compositional score matrix adjustment) and PSI-BLAST (*E*-value cutoff of 0.005). BoNT A1 (NCBI gi: 152932838) and NTNH A1 (NCBI gi: 152931876) were used as representative CNT queries.

PFAM domains were predicted using hmmsearch [HMMER 3.1b1 package [10]], using the following domain models: Peptidase M27 (PF01742), translocase (PF07952), N-terminal binding domain (PF07953), C-terminal binding domain (PF07951), and the NTNH C-terminal binding domain (PF08470). Heuristic filters were turned off for maximum sensitivity using the max parameter.

4.2. Alignment and phylogenetic analysis

Representative BoNT sequences from each toxinotype and TeNT were retrieved partly based on Dover et al. [19] along with their neighboring NTNH genes. An alignment of these sequences

including ORF1 and ORF2 was constructed with MUSCLE v3.8.31 [29]. ORF2 was truncated (1–495 aa) to exclude non-homologous regions.

A maximum likelihood phylogenetic tree was constructed using RAxML [30] version 8.0.26, with the FLU + GAMMA model of evolution (chosen by automated model selection), and 100 rapid bootstrap replicates. A neighbour joining (NJ) tree was also constructed using the BioNJ algorithm as implemented in SeaView [31] and Poisson distances (gaps included), with 100 bootstrap replicates. The alignment, trees and accession numbers are available in Supplementary information.

For 16S small subunit analysis, prealigned *C. botulinum* and *Weissella* sequences were retrieved from the SILVA (http://www.arb-silva.de) database, and percentage identities were calculated using the alistat program within the EMBOSS package (http://emboss.open-bio.org).

5. Accession numbers

NCBI gi numbers of CNTs, ORF1 and ORF2 used in the paper:

BoNT/A1 NCBI gi: 152932838, BoNT/A2 NCBI gi: 226841116, BoNT/A3 NCBI gi: 169409115, BoNT/BvA4 NCBI gi: 229260380, BoNT/A5 NCBI gi: 253584711, BoNT/B1 NCBI gi: 169123155, BoNT/B2 NCBI gi: 670451768, BoNT/B3 NCBI gi: 1123229145, BoNT/B4 NCBI gi: 187723989, BoNT/BvB NCBI gi: 229260169, BoNT/B6 NCBI gi: 159132570, BoNT/B7 NCBI gi: 344939560, BoNT/C NCBI gi: 169294107, BoNT/C/D NCBI gi: 658092732, BoNT/D NCBI gi: 658083582, BoNT/D/C NCBI gi: 253960377, BoNT/E1 NCBI gi: 243083737, BoNT/E2 NCBI gi: 123229171, BoNT/E3 NCBI gi: 188500413, BoNT/E4 NCBI gi: 237655392, BoNT/E5 NCBI gi: 9651019, BoNT/E6 NCBI gi: 160419788, BoNT/E7 NCBI gi: 353558547, BoNT/E8 NCBI gi: 353558549, BoNT/E9 NCBI gi: 410825131, BoNT/F1 NCBI gi: 152935704, BoNT/F2 NCBI gi: 282160555, BoNT/F3 NCBI gi: 282160573, BoNT/F4 NCBI gi: 282160565, BoNT/F5 NCBI gi: 282160587, BoNT/F6 NCBI gi: 316925157, BoNT/F7 NCBI gi: 525547321, BoNT/G NCBI gi: 441276, TeNT NCBI gi: 28208772, NTNH/A1 NCBI gi: 152931876, NTNH/A2 NCBI gi: 226843051, NTNH/A3 NCBI gi: 169409021, NTNH/BvA4 NCBI gi: 229260305, NTNH/A5 NCBI gi: 253584710, NTNH/B1 NCBI gi: 169123249, NTNH/B2 NCBI gi: 670451767, NTNH/B4 NCBI gi: 187723959, NTNH/BvB NCBI gi: 229260237, NTNH/B6 NCBI gi: 159132569, NTNH/C NCBI gi: 169294113, NTNH/C/D NCBI gi: 658092731, NTNH/D NCBI gi: 658083581, NTNH/E1 NCBI gi: 243083712, NTNH/E3 NCBI gi: 188499467, NTNH/E4 NCBI gi: 237657246, NTNH/E6 NCBI gi: 160419787, NTNH/F1 NCBI gi: 152934837, NTNH/F6 NCBI gi: 316925156, NTNH/F7 NCBI gi: 525547321, NTNH/G NCBI gi: 2104804, ORF1 NCBI gi: 653854119, ORF2 NCBI gi: 653854116.

6. Other accession numbers

W. oryzae genome: NCBI Genome ID 31977, *C. botulinum* str. A Hall and *C. botulinum* str. F Langeland NCBI Genome ID: 726, *W. koreensis* KACC 15510 NCBI Genome ID: 6922, *W. oryzae* contig (NZ_BAWR01000013.1), *L. mesenteroides* ATCC 8293 (NC_008531.1), *W. koreensis* KACC 15510 (NC_015759.1).

Acknowledgment

This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC), through a Discovery Grant to A.C.D.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2014.12.018.

References

[1] Montecucco, C. and Molgó, J. (2005) Botulinal neurotoxins: revival of an old killer. Curr. Opin. Pharmacol. 5, 274–279.
[2] Peck, M.W., Stringer, S. and Carter, A.T. (2011) *Clostridium botulinum* in the post-genomic era. Food Microbiol. 28, 183–191.
[3] Hill, K.K. and Smith, T. (2013) Genetic diversity within *Clostridium botulinum* serotypes, botulinum neurotoxin gene clusters and toxin subtypes. Curr. Top. Microbiol. Immunol. 364, 1–20.
[4] Montal, M. (2010) Botulinum neurotoxin: a marvel of protein design. Annu. Rev. Biochem. 79, 591–617.
[5] Rossetto, O., Pirazzini, M. and Montecucco, C. (2014) Botulinum neurotoxins: genetic, structural, and mechanistic insights. Nat. Rev. Microbiol. 12, 535–549.
[6] Humeau, Y., Doussau, F., Grant, N.J. and Poulain, B. (2000) How botulinum and tetanus neurotoxins block neurotransmitter release. Biochimie 82, 427–446.
[7] Sharma, S.K., Ramzan, M.A. and Singh, B.R. (2003) Separation of the components of type A botulinum neurotoxin complex by electrophoresis. Toxicon 41, 321–331.
[8] Doxey, A., Lynch, M., Müller, K., Meiering, E. and McConkey, B. (2008) Insights into the evolutionary origins of clostridial neurotoxins from analysis of the *Clostridium botulinum* strain A neurotoxin gene cluster. BMC Evol. Biol. 8, 316.
[9] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[10] Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.
[11] Tanizawa, Y., Fujisawa, T., Mochizuki, T., Kaminuma, E., Suzuki, Y., Nakamura, Y. and Tohno, M. (2014) Draft genome sequence of *Weissella oryzae* SG25T, isolated from fermented rice grains. Genome Announc. 2. e00667-14.
[12] Guy, L., Kultima, J.R. and Andersson, S.G. (2010) GenoPlotR: comparative gene and genome visualization in R. Bioinformatics 26, 2334–2335.
[13] Tohno, M., Kitahara, M., Inoue, H., Uegaki, R., Irisawa, T., Ohkuma, M. and Tajima, K. (2013) *Weissella oryzae* sp. nov., isolated from fermented rice grains. Int. J. Syst. Evol. Microbiol. 63, 1417–1420.
[14] Sobel, J., Tucker, N., Sulka, A., McLaughlin, J. and Maslanka, S. (2004) Foodborne botulism in the United States, 1990–2000. Emerg. Infect. Dis. 10, 1606–1611.
[15] Lee, J.S., Lee, K.C., Ahn, J.S., Mheen, T.I., Pyun, Y.R. and Park, Y.H. (2002) *Weissella koreensis* sp. nov., isolated from kimchi. Int. J. Syst. Evol. Microbiol. 52, 1257–1261.
[16] Björkroth, K.J., Schillinger, U., Geisen, R., Weiss, N., Hoste, B., Holzapfel, W.H., Korkeala, H.J. and Vandamme, P. (2002) Taxonomic study of *Weissella confusa* and description of *Weissella cibaria* sp. nov., detected in food and clinical samples. Int. J. Syst. Evol. Microbiol. 52, 141–148.
[17] Magnusson, J., Jonsson, H., Schnürer, J. and Roos, S. (2002) *Weissella soli* sp. nov., a lactic acid bacterium isolated from soil. Int. J. Syst. Evol. Microbiol. 52, 831–834.
[18] Niemann, H., Schubert, W. and Heinz, D. (2004) Adhesins and invasins of pathogenic bacteria: a structural view. Microbes Infect. 6, 101–112.
[19] Dover, N., Barash, J.R., Hill, K.K., Xie, G. and Arnon, S.S. (2013) Molecular characterization of a novel botulinum neurotoxin type H gene. J. Infect. Dis. 209, 192–202.
[20] Hill, K.K., Xie, G., Foley, B., Smith, T.J., Munk, A.C., Bruce, D., Smith, L.A., Brettin, T.S. and Detter, J.C. (2009) Recombination and insertion events involving the botulinum neurotoxin complex genes in *Clostridium botulinum* types A, B, E and F and *Clostridium butyricum* type E strains. BMC Biol. 7, 66.
[21] Carter, A.T., Peck, M.W. (2014). Genomes, neurotoxins and biology of Clostridium botulinum Group I and Group II. Res. Microbiol., http://dx.doi.org/10.1016/j.resmic.2014.10.010 (Epub ahead of print).
[22] Carter, A.T., Stringer, S.C., Webb, M.D. and Peck, M.W. (2013) The type F6 neurotoxin gene cluster locus of group II *Clostridium botulinum* has evolved by successive disruption of two different ancestral precursors. Genome Biol. Evol. 5, 1032–1037.
[23] Park, J.B. and Simpson, L.L. (2003) Inhalational poisoning by botulinum toxin and inhalation vaccination with its heavy-chain component. Infect. Immun. 71, 1147–1154.
[24] Costa, F.A.A., Leal, C.A.G., Schuenker, N.D., Leite, R.C. and Figueiredo, H.C.P. (2014) Characterization of *Weissella ceti* infections in Brazilian rainbow trout, *Oncorhynchus mykiss* (Walbaum), farms and development of an oil-adjuvanted vaccine. J. Fish Dis., http://dx.doi.org/10.1111/jfd.12236 (Epub ahead of print).
[25] Fairfax, M.R., Lephart, P.R. and Salimnia, H. (2014) *Weissella confusa*: problems with identification of an opportunistic pathogen that has been found in fermented foods and proposed as a probiotic. Front. Microbiol. 5, 254.
[26] Nieminen, T.T., Säde, E., Endo, A., Johansson, P. and Björkroth, J. (2014) The family *Leuconostocaceae*. Prokaryotes, 215–240.
[27] DasGupta, B.R. (2006) Botulinum neurotoxins: perspective on their existence and as polyproteins harboring viral proteases. J. Gen. Appl. Microbiol. 52, 1–8.

[28] Lipka, V., Kwon, C. and Panstruga, R. (2007) SNARE-Ware: the role of SNARE-domain proteins in plant biology. Annu. Rev. Cell Dev. Biol. 23, 147–174.
[29] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
[30] Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
[31] Gouy, M., Guindon, S. and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.